# Psychophysical Comparisons of Single- and Dual-band Fused Imagery

Michael J. Sinai<sup>1</sup>, Jason S. McCarley<sup>1</sup>, William K. Krebs<sup>1</sup>, Edward A. Essock<sup>2</sup>

<sup>1</sup>Naval Postgraduate School, Monterey, CA <sup>2</sup>University of Louisville, Louisville, KY

#### ABSTRACT

The purpose of this study was to investigate whether a dual-band sensor fused image improves visual performance compared to a single-band image. Specifically, we compared behavioral performance using images from an uncooled LIMIRS long-wave infrared sensor and a Fairchild image intensified low-light CCD, against these same images after they had been 'fused' by combining both spectral bands into a two-dimensional color space <sup>1</sup>. Human performance was assessed in two experiments. The first experiment required observers to detect target objects presented against naturalistic backgrounds and then identify whether those detected targets were vehicles or persons. The second experiment measured observers' situational awareness by asking them to rapidly discern whether an image was upright or inverted. Performance in both tasks, as measured by reaction times and error rates, was generally best with the sensor-fused images, although in some instances performance with the single band images was as good as performance using the sensor-fused images. Results suggest that sensor fusion may facilitate human performance both by facilitating target detection and recognition, and by enabling higher levels of more general situational awareness and scene comprehension.

**Keywords:** Sensor fusion, multi-spectral displays, infrared, human performance

## 1. INTRODUCTION

The prospect of multi-band sensor fusion as an aid to human performance has provoked a great deal of scientific interest in recent years. By combining information from multiple single-band sources within a unitary display, researchers hope to overcome perceptual limitations inherent in the images provided by various electro-optical sensors singly. Of particular interest is the possibility of fusing imagery provided by low-light image intensified (LL) and long-wave infrared (IR) sensors, those currently used to support night military operations. These two types of sensors offer complimentary information about a scene, one from intensified reflected visible light, and the other from thermal emissivity. The fusion of these two different image types involves localized comparisons of these two images, either in a center/surround-like arrangement modeled after biological visual processing <sup>2</sup>, or by pixel by pixel comparisons <sup>1,3,4,5</sup>. Because these sensors respond to widely separated wavebands within the electromagnetic spectrum (LL sensors to long-wave visible near infrared energy, IR sensors to emitted thermal radiation), each maintains advantages and suffers disadvantages that the other does not, which can change depending on the atmospheric and environmental conditions. For example, the resolution on most infrared sensors is generally poorer than image intensified sensors. However, the contrast between heat emitting objects and their surroundings is greater in the infrared scene compared to the image-intensified scene. Also, different atmospheric conditions will affect these two sensors differently. For example, clouds can obscure moonlight and starlight, which would adversely affect the image intensified scenes but would not affect the infrared scenes. Likewise different times of the day will affect the heat emitted from various objects within a scene but would not affect the scene from an image-intensified sensor, except of course unless the ambient illumination also changed. Accordingly, it is hypothesized that by fusing these two complimentary sensors into a single false color scene may result in equal to or better operator target detection performance compared to the two single band sensors alone.

The aim of dual-band sensor fusion is to create a composite image that simultaneously presents information derived from two sensors, thereby allowing a human operator user to utilize both sensors without having to attend to two displays. Another potential advantage afforded by sensor fusion is to provide scene information not present in either single band image alone by deriving information based on the differences between the sensors. Furthermore, many sensor fusion algorithms <sup>1, 2, 6, 7</sup> add color to the composite image, providing even more potentially useful information about the scene. However, despite the benefits that are commonly assumed with sensor fusion, the experimental evidence in support of sensor fusion has been equivocal. While some studies have found performance was better with fused imagery <sup>2, 4, 8, 9</sup>, others have not <sup>10, 11, 12</sup>. Part of the discrepancy of these studies stems from the different fusion algorithms employed as well as the wide range of tasks used to measure behavior. While many of these studies have focused on object detection tasks, the procedures for testing have varied dramatically, ranging from testing small patches of the original scenes that were briefly flashed <sup>9, 12</sup>, to changing the contrast of a small square embedded in the scene <sup>2</sup>, to using video clips of the processed imagery <sup>11</sup>. It is not surprising these very different methodologies found different results. Past research then, has been somewhat inconclusive in determining under what conditions and for what task sensor fusion improves human performance.

The current study addressed this issue by comparing performance on two different tasks, an object recognition task and a situational awareness task. The first task measured observers' ability to identify an object imbedded within a natural nighttime scene. The target objects were people and vehicles with some trials containing no targets - false alarm trials. Thus, the observer's task was not only to detect the presence of an object, but also they determine whether the object was a person or a vehicle. The second task measured subjects global processing of the entire image by requiring the observer to determine whether the scene was inverted or upright. It was hypothesized that performance on these two very different tasks would be differently affected both by the single sensor imagery and by the fused imagery. That is, the infrared image usually has higher contrast than the image-intensified image, we hypothesized performance on the detection/recognition task may be somewhat better for the infrared imagery compared to the imageintensified imagery. Likewise, because the infrared imagery has lower resolution than the imageintensified imagery, we hypothesized that performance would be slightly better for the image-intensified imagery compared to the infrared imagery for the situational awareness task because of the need for scene detail in such a task. Furthermore we hypothesized that the fused imagery would result in performance at least as good as the better of the two single band sensors. These predictions were generally supported by the results.

# 2.0 METHODS

# 2.1 Subjects

Sixty active duty military personnel were recruited for this experiment. All had normal or corrected to normal acuity by self report, and all signed informed consent forms prior to participation

# 2.2 Apparatus

Stimuli were displayed by a VisionWorks computer graphics system (Vision Research Graphics, Inc., Durham New Hampshire)  $^{13}$  on a Nanao Flexscan F2.21 monitor. The monitor had a resolution of 800 x 600 pixels, a frame rate of 98.9 Hz, and a maximum luminance of 100 cd/m $^2$  with luminance linearized by means of a look up table. Observers viewed the screen from 1.5 meters.

# 2.3 Stimuli

Stimuli were images collected at Fort AP Hill, VA (.04 lux) using an uncooled LIMIRS long-wave infrared sensor (IR) and a Fairchild image intensified low-light CCD (LL). The images were of various nighttime scenes from around the installation, including wooded areas, fields, roads, and buildings. Images were then spatially registered and 'fused' by combining both spectral bands into a two-dimensional color space through an algorithm developed at the Naval Research Laboratory 1,6. The fusion algorithm used assigns each pixel a color vector determined by the detected power in the registered LL and IR imagery, with pixels that differ in their values of combined IR and LL power being presented in different intensities, and pixels that differ in their ordinal emissivity and reflectivity being presented in different hues. From a principal components analysis of the set of pixels in IR/LL space, a principal component direction is taken to correspond approximately to an illuminant/radiant intensity vector 1. Intensity is assigned to the correlated component (major axis) for each pixel, and color is assigned to the uncorrelated feature (minor axis). Pixel color is assigned by opposing LL intensity against IR intensity and assigning one hue (cyan) to pixels whose intensity is greater in the LL than the IR image, and another hue (red) to pixels whose intensity is greater in the IR than in the LL image. Thus, the resulting image is displayed in two hues of various saturations. This coding system produces false-color imagery in which hue directly indicates the ordinal relative intensities of emitted and reflected energy at each pixel, lending a potential advantage for some perceptual tasks <sup>1</sup>.

A total of six image formats were tested: single-band IR and LL formats, two color-fused formats, and two achromatic fused formats. One color-fused version of each scene was derived using IR imagery of white-hot polarity, and the other using IR imagery of black-hot polarity. Achromatic versions of these fused images were spatially identical to their chromatic counterparts, but were rendered in grayscale. Single-band IR was of white-hot polarity. A total of sixty-six scenes were used in the two experiments, not counting practice scenes. Each scene was rendered in the 6 different formats. All images had dimensions of  $625 \times 400$  pixels. The surrounding screen was kept at a constant  $50 \text{ cd/m}^2$  throughout experiment.

### 2.4 Procedure

Observers participated in both experiments with order in which the experiment was presented counterbalanced across subjects. Each observer viewed imagery from only one sensor format for both tasks, and each scene was used only once during all experiments. A randomized block design was used to determine which format each subject viewed.

For the object identification task, each trial began with an auditory tone followed immediately by the image displayed on the screen. The observer's task was to decide whether a person, vehicle, or neither were present anywhere in the scene. Responses were indicated with a keypress ("1" to indicate a person present, "2" to indicate a vehicle present, and "3" to indicate neither were present). The scene remained visible until a response was made, which then initiated the next trial. No feedback was given during the experiment. Twelve practice trials were followed by thirty-six experimental trials. Not counting the practice trials, twelve scenes contained people, twelve contained vehicles, and twelve contained neither. Accuracy and reaction time were recorded.

The procedure was the same for the situational awareness task, except the task was for the observers to decide whether the scene was inverted or not. Responses were indicated with a keypress ("1" to indicate the scene was upright, "2" to indicate the scene was inverted). Ten practice trials were followed by 30 experimental trials, half of which were randomly inverted.

#### 3.0 RESULTS

Reaction times for incorrect responses were discarded. RTs and error rates from both experiments were submitted to separate 6 x 3 ANOVA's with image format and target type as factors. Mean RTs for target identification are presented in Figure 1. Analysis of RTs for the identification task did not reveal a significant main effect of format,  $(F_{(5,54)} = 1.08, p = .38)$ , however there was a significant main effect of

target type  $(F_{(2,\ 108)}=48.7,\ p<.0001)$ . The interaction between image format type and target was not significant  $(F_{(10,\ 108)}=.41,\ p=.94)$ . Figure 2 shows the data collapsed over target type for each of the format types. While the reaction times are slightly longer for the IR and LL conditions, the difference did not reach statistical significance. Mean error rates for target identification are shown Figure 3. The ANOVA for error rates on the identification task found a significant effect for format type  $(F_{(5,\ 54)}=3.34,\ p<.01)$ , target type  $(F_{(2,\ 108)}=13.59,\ p<.0001)$ , and the interaction  $(F_{(10,\ 108)}=2.12,\ p<.03)$ . Figure 4 shows the same results when averaged over target type and plotted by image format type. Tukey's post-hoc t test revealed that the color fused white hot format type had significantly lower error rates than its grayscale counterpart.

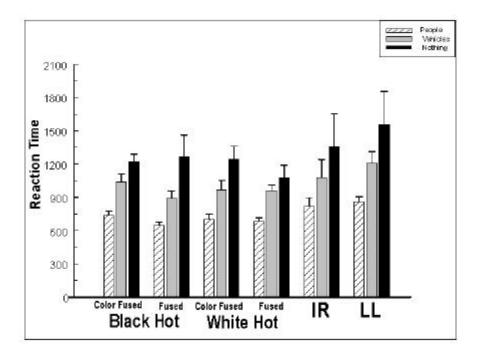
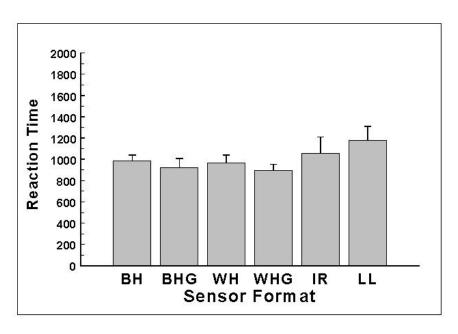


Figure 1. Mean reaction times for the identification task. Error bars represent 1 SEM.



**Figure 2**. Mean reaction times for the identification task averaged across target types and plotted by format type. Error bars represent 1 SEM. Bh- black hot color fused format, bhg- black hot grayscale fused, wh – white hot color fused format, IR- infrared format, LL- low light image intensified format.

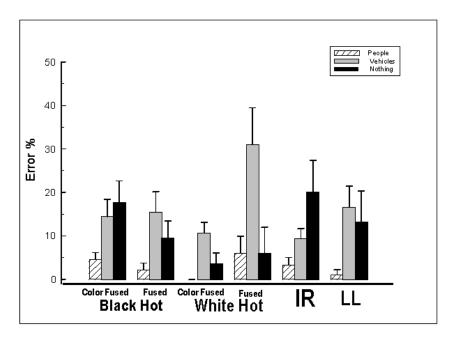
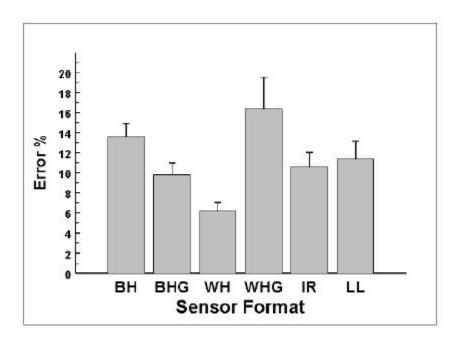


Figure 3. Error rates for the identification task. Error bars represent 1 SEM.



**Figure 4**. Mean error rates for the identification task averaged across target types and plotted by format type. Error bars represent 1 SEM.

The ANOVAs for the situational awareness task showed a significant effect for image format type for reaction times ( $F_{(5,54)} = 5.93$ , p <.001), and error rates ( $F_{(5,54)} = 4.66$ , p <.002). These results are shown in Figures 5 and 6. Post-hoc t tests found only the IR condition was significantly slower than the other formats for reaction times, and that the IR condition as well as the white hot grayscale fused were significantly worse than the other conditions for error rates.

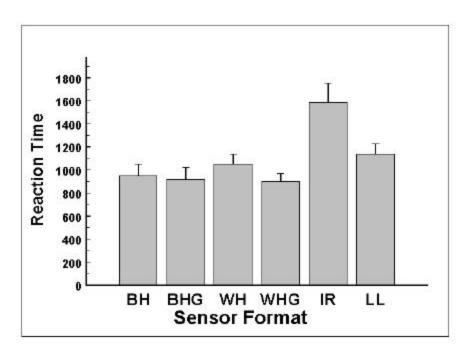


Figure 5. Mean reaction times for the situational awareness task

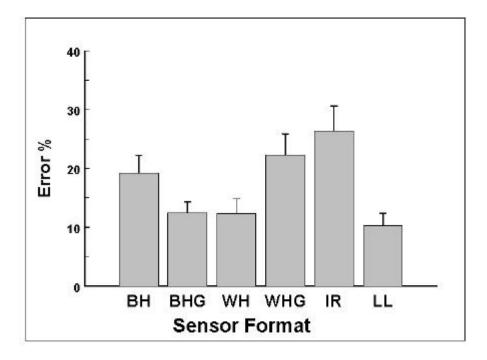


Figure 6. Mean error rates for the situational awareness task

# 4.0 DISCUSSION

Significant differences were found between single-and dual-band image format types for two psychophysical tasks, an object identification task and a situational awareness task. Even though error rates were lower for the white-hot color fused format compared to the single-band formats in the object ID task,

the largest differences occurred between the white-hot color fused and the white-hot grayscale fused formats. This large difference between the color and grayscale version of the same format underscores the advantage afforded by the addition of color to these images for this task. For the situational awareness task significant differences were found between the single-band and dual band formats for both error rates and reaction times. In both cases, the IR format was worse than the other conditions and for error rates the white-hot fused grayscale also was significantly worse than the other conditions. In this task, the IR format was both slower and more error prone than the other formats. This negative effect may have been due to the lower resolution of the IR imagery which degraded the spatial content of the scene.

The effects of color were very strong in this study. Significant differences were found between the white-hot color fused error rates and the white-hot grayscale fused error rates for both tasks. Thus the false-color of the fusion algorithm improved performance for this format type both when identifying objects among a cluttered background and when determining the orientation of the scene. The only difference between the two formats was the addition of color. Other studies have also found the addition of false color to the fused imagery improves performance in some conditions <sup>4, 9, 12</sup>. In the other fused format tested, however, we did not find that color improved performance. For the black-hot fused format, the addition of color information to the images actually hindered performance by increasing error rates in both tasks. However the deleterious effects of color for the black-hot fused format was not as much as the benefits of color to the white-hot format. The only difference between the two fused formats tested was the polarity of the IR input, whether it was white hot or black hot. The fact that this small difference in the input had such a large difference in performance underscores the sensitive nature of the fusion process and the need for specific behavioral testing. The results from this study provide compelling evidence for the benefits in performance that can be found using color-fused imagery.

#### 6.0 ACKNOWLEDGMENTS

Support for this research was provided by Office of Naval Research grants #N0001497WR30078, #N000149WR30091, #N0001498AF00002, DARPA grant #N0017398WR00301, and by the Lockheed Martin Corporation University Research Grant program.

We thank Dr. Dean Scribner, Dr. Penny Warren, and Mr. Jon Schuler for their efforts in processing the still frame images.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the United States Government.

#### 7.0 REFERENCES

- 1) Scribner, D.A., Satyshur, M.P., Schuler, J. and Kruer, M.P. (1996). Infrared color vision. <u>IRIS Specialty Group on Targets, Backgrounds and Discrimination</u>, January, Monterey, CA.
- 2) Waxman, A.M., Gove, A.N., Seibert, M.C., Fay, D.A., Carrick, J.E., Racamato, J.P., Savoye, E.D., Burke, B.E., Reich, R.K., McGonagle, W.H. and Craig, D.M. (1996). Progress on color night vision: Visible/IR fusion, perception and search, and low-light CCD imaging. <u>Proceedings of the SPIE Conference on Enhanced and Synthetic Vision, SPIE-2736</u>, 96-107.
- 3) Toet, A. & Walraven, J. (1996) New false color mapping for image fusion. <u>Optical Engineering</u>, <u>35</u>, 650-658.
- 4) Toet, A., Ijspeert, J. K., Waxman, A. M., and Aguilar, M. (1997) Fusion of visible and thermal imagery improves situational awareness. <u>Proceedings of the SPIE Conference on Enhanced and Synthetic</u> Vision 1997, SPIE-3088, 177-188.

- 5) Therrien, C.W., Scrofani, J., and Krebs, W.K. (1997). An adaptive technique for the enhanced fusion of low-light visible with uncooled thermal infrared imagery. <u>Proceedings of the IEEE: International</u> Conference on Imaging Processing, pp. 405-408, October 1997.
- 6) Scribner, D.A., Warren, P., Schuler, J. Satyshur, M.P., and Kruer, M.P. (1998). Infrared color vision: an approach to sensor fusion. Optics and Photonics News, 8, 27-32.
- 7) Krebs, W.K., McCarley, J.M., Kozek, T., Miller, G.M., Sinai, M.S., and Werblin, F.S. (1999) An evaluation of a sensor fusion system to improve drivers' nighttime detection of road hazards. <u>Investigative Opthalmology and Visaul Sciences, Suppl.</u>, in press.
- 8) Sinai, M.S, McCarley, J.S., and Krebs, W.K. (1999). A comparison of sensor fusion and single band sensors in the recognition of nighttime scenes. <u>Proceedings of the 1999 Meeting of the IRIS Specialty Group on Passive Sensors</u>, Monterey, CA.
- 9) Essock, E.A., Sinai, M.J., McCarley, J.S. and Krebs, W.K., and DeFord, J. K. (1999a) Perceptual ability with real-world nighttime scenes: Image-intensified, infrared and fused-color imagery. <u>Human Factors</u>, in press.
- 10) Steele, P.M. and Perconti, P. (1997). Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage. <u>SPIE 11th Annual International Symposium on Aerospace/Defense Sensing, Simulation and Controls, Orlando FL.</u>
- 11) Krebs, W.K., Scribner, D.A., Miller, G.M., Ogawa, J.S., Schuler, J. (1998). Beyond third generation: a sensor fusion targeting FLIR pod for the F/A-18. <u>Proceedings of the SPIE-Sensor Fusion: Architectures, Algorithms, and Applications II, 3376</u>, 129-140.
- 12) Essock, E.A., Krebs, W.K., Sinai, M.J., DeFord, J. K., Srinivasan, N., and McCarley, J.S.(1999b) Human Perceptual Performance With Nighttime Imagery: Region Recognition and Texture-Based Segmentation. Manuscript in preparation.
- 13) Swift, D.J., Panish, S. and Hippensteel, B. (1997) The use of VisionWorks in visual psychophysics research. Spatial Vision, 10, 471-477.

Further author information -

W.K.K. (correspondence): Email: wkrebs@nps.navy.mil Telephone: (831) 656-2543; Fax: (831) 656-2595